基于音频基频特征的鲁棒零水印算法 *

郑梦怡，李　晨，田丽华†

(西安交通大学 软件学院, 西安 710049)

摘　要：为了将数字音频水印技术与音频内容相结合，基于人声和乐器的基频稳定的特点，从音乐中提取出基频并据此设计一种鲁棒零水印方法。首先通过正规化分谐波叠加算法提取基频，然后采用 K-means 算法对基频特征进行编码加强其稳定性，最后与水印图像异或生成零水印序列。此外，发生恶意窜改时，对比零水印中的基频信息得到不一致的部分，即可确定恶意窜改区域。实验结果表明，此算法在常规攻击及抖动攻下均具有良好的稳定性，且可实现窜改检测。

关键词：数字音频水印；鲁棒零水印；正规化分谐波叠加；基频；窜改检测

中图分类号：TP391　　doi: 10.3969/j.issn.1001-3695.2017.11.0942

Robust zero-watermarking algorithm based on audio fundamental frequency feature

Zheng Mengyi, Li Chen, Tian Lihua†

(*School of Software Engineering*, *Xi'an Jiaotong University*, *Xi'an* 710049, *China*)

Abstract: In order to combine the digital audio watermark technology with the audio content, because of the stability of the fundamental frequency of the human voice and musical instruments, this paper designed a robust audio zero-watermarking method based on the fundamental frequency of music. Firstly, the normalized subharmonic summation algorithm extracted the fundamental frequency from each frame of the audio signal. Then, the K-means algorithm encoded the extracted fundamental frequencies so as to further improve the stability of the feature. Finally, the encoded fundamental frequencies XOR with the watermark image to generate a zero watermark sequence. Besides, when the audio is maliciously tampered, it can get the inconsistent part of the audio signal by comparing the fundamental frequency information contained in the zero watermark sequence. At the inconsistent section, the algorithm can detect the maliciously tampered area. The experiment results show that the proposed algorithm has good robustness against common attacks and jitter attacks, and the results also prove that the algorithm can detect and locate the tampering.

Key Words: digital audio watermarking; robust zero watermark; normalized subharmonic summation; fundamental frequency; tamper detection

0　引 言

极速发展的计算机和互联网络技术促使数字产品的产生、保存和传输变得更加便利。与此同时，非法复制等盗版问题也越来越严重。为了解决数字产品的恶意侵权及版权纠纷问题，水印技术作为一种版权保护技术被提出，并且已经成为当前解决此类问题的有效工具[1]。

数字音频水印技术是在保证听觉效果的基础上，把水印信息隐藏于音频中的技术，实现对数字音频产品的版权保护，即嵌入水印后保证音频具有透明性[2]。文献［3］根据 DCT 系数符号具有稳定性的特点提出一种零水印算法，该算法鲁棒性较好，但是作为音频特征的 DCT 系数符号不备实际意义。文献［4］提出一种基于 DWT-SVD 的零水印算法，该算法需要保证音频长度大于 1 024 倍的水印图像尺寸，且音频大于所需部分被舍弃，对测试音频的选择有严格要求；此外，对于抖动攻击该算法鲁棒性不强。文献［5］所述的 DWT 与 SVD 相结合的水印算法同文献[4]思路相似，虽然对于普通攻击鲁棒性较好，然而同文献［4］存在的问题类似，提取的特征不能直观代表版权信息。文献［6］提出一种基于音频特征和低频系数较小值的水印算法，该算法需要选取适当的阈值舍弃部分音频帧，算法的稳定性欠佳，且对于 MP3 压缩攻击鲁棒性较差。文献［7］提出用于内容认证的零水印算法，构造零水印的特征是音频低频分量，该算法对于抵抗滤波攻鲁棒性较弱。文献［8］提出一种基于音频帧的过零率和短时能量特征的水印算法，该算法简

基金项目：国家自然科学基金资助项目（61403302）；西安交通大学科研业务基金项目（XJJ2016029）

作者简介：郑梦怡（1991-），女，河南南阳人，硕士研究生，主要研究方向为数字音频水印；李晨（1981-），女，陕西西安人，讲师，博士，主要研究方向为多媒体技术；田丽华（1978-），女（通信作者），陕西西安人，高级工程师，博士，主要研究方向为多媒体技术（lhtian@mail.xjtu.edu.cn）.

单易懂，但是对于噪声和 MP3 压缩攻击鲁棒性较差。

由于音频中基频特征具有稳定性，且作为音频特征与各种变换系数相比具有直观的实际意义，本文提出了一种基于音频基频特征的鲁棒零水印算法。首先对输入音频信号进行小波近似系数重构预处理；然后根据正规化分谐波叠加算法计算每帧的多个基频，根据候选基频集提取能量最强的基频。由于信号的基频具有比较稳定的特点，所以该特征生成的零水印序列抵抗常见攻击具有良好的鲁棒性。由于各帧的基频特征在提取过程中互不影响，当音频遭到局部窜改时，通过比较窜改前后的基频特征序列，可以确定特征改变的帧，这些发生变化的帧所在的区域即为窜改区域，所以本算法可以实现对恶意窜改的准确定位。

1　理论基础

1.1　零水印

零水印技术是一种无须实际向载体信号中嵌入水印信息的水印技术，是根据音频信号的某些重要特征，将特征信息与水印图像结合生成"零水印"[9]。因此，零水印算法不存在嵌水印过程中产生的隐蔽性与鲁棒性之间冲突的问题。

1.2　音乐基频特征

音乐的基频是音频振动系统中频率最低的固有成分（即基音）处的频率，它决定音高，作为音频信号的最基本特征之一，在音乐信息检索系统中被广泛使用[10]。而本文选取基频作为音频信号特征的主要原因是基频具有稳定性且可表达音频内容信息的特性。

1.3　正规化分谐波叠加（NSHS）算法

基于音频的谐波位于基频的倍频处的特点，分谐波叠加(shs)算法通过把各次谐波的能量不断叠加到基频处，实现增强基频能量突出基音的目的[11]。Hsu 等人[12]提出的正规化分谐波叠加算法，是在传统分谐波叠加算法的基础上，把每个频率乘以一个系数$1/\sum_{n=1}^{N_f} h_n$,避免了传统算法因多次叠加造成低频能量的过度增加的缺点，达到在不过度增加低频及基频能量的前提下，削减高阶谐波对基音产生的影响，准确提取出基频的目的[13,14]。其中叠加谱定义如下：

$$H(f)=\frac{\sum_{n=1}^{N_f} h_n P_t(nf)}{\sum_{n=1}^{N_f} h_n} \tag{1}$$

其中：$h_n=0.84^{n-1}$；$N_f=floor\left(\frac{f_s}{2f}\right)$；$f_s$为采样频率。

2　音频基频特征提取

算法提取基频特征的主要工作是首先得到每帧音频信号的多个基频，再在每一帧的基频候选集中选取能量最大的基频作为此帧的特征基频，这些基频不一定属于音频中的同一声源，每帧提取出的基频特征不需要根据某种声源进行筛选，一方面是为了避免由于某些帧某种声源的停顿导致此帧出现特征缺失的情况，另一方面由于每帧的基频仅根据当前帧能量分布情况提取，无须根据某种特定声源的限制进行基频特征的选择取舍，所以彼此间没有直接关联，当某一帧提取出的基频特征发生改动时不会影响其余帧基频的提取。因此，本方案得到的音频基频特征序列不仅具有基频特征固有的稳定性的特点，还可以用于窜改检测。图 1 描述了特征提取的基本流程。

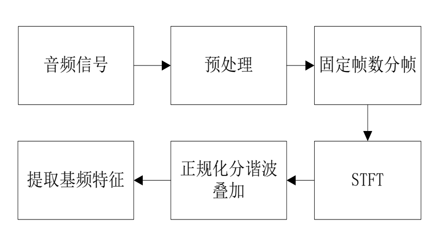


图 1　音频基频提取流程

2.1　预处理

由于人声和乐器的基频通常位于中低频，为降低高频信息对提取基频特征的影响，首先对原始音频进行小波变换近似系数重构的预处理。近似系数就是信号中大尺度的低频部分，很多信号中低频信息是最重要的部分，基本包含了信号的基本特征。图 2(上)所示的波形图为原始音频，图 2(下)显示的是 3 级小波变换近似系数重构后的音频波形。

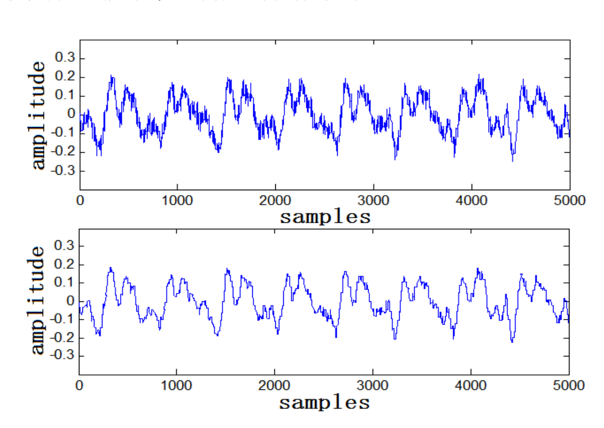


图 2　原始音频信号与小波变换近似系数重构信号

由图 2 可以看出，小波变换近似系数重构后的音频信号比较光滑，高频部分得到抑制，一定程度上降低了噪声影响，有利于基频特征的提取。

2.2　固定帧数分帧

音频预处理后，对其进行固定帧数有重叠的分帧。

采用有重叠的分帧方法的原因是音频信号的变化是连续的，所以分帧后音频信号的特征参数应该平滑变化，通过有重叠的分帧可以使相邻帧间首尾衔接平缓过渡，避免提取出的特征参数不连续。

这里选择固定帧数分帧是为了减少时间尺度上的伸缩对提取基频特征造成的影响。其次，由于构造零水印需要从每帧提取出一个特征点与一位水印信息进行异或运算，特征点的个数

要与待嵌入的水印图像的像素点数相一致，所以分帧的个数由水印图像确定为一个固定的值。

2.3 提取基频特征

对分帧后每一帧音频信号进行 STFT 转换到频域，然后采用 NSHS 算法提取每帧的候选基频集。由于根据 NSHS 算法，各次谐波被搬移到基频上并对基频能量进行叠加，所以基频处的能量相对于其他频率处的能量得到增强，基频处的能量峰值更突出，有益于提取基频特征。图 3(上)是原始音频某一帧的频谱图，图 3(下)是同一帧正规化分谐波叠加后的频谱图。

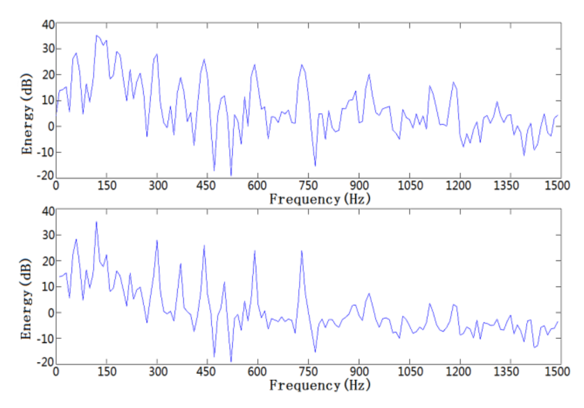


图 3　原始音频频谱及 NSHS 叠加结果

由图 3 可以看出，经过 NSHS 叠加后，在基频处(约 150 Hz)出现了明显的峰值，非峰值部分比较平稳，并且高频部分的能量明显得到抑制，由于局部能量峰值处的频率构成所求候选基频集，所以对谐波叠加后的信号进行峰值检测及提取候选基频更加有利。同时，由图 3 可以看出，叠加后谐波及其他频率处的能量均低于基频能量，所以，对于 NSHS 叠加后的每帧信号提取出的候选基频集，能量最强的频率即为这一帧的基频特征，将其加入到基频特征序列中并保存。

图 4 为各种攻击下提取出的基频特征 F0 与原始音频基频特征 F0 的对比结果。图中，实线表示原始音频基频特征，虚线代表遭到攻击后的音频提取的基频特征；fs 为音频原始采样频率。若攻击后的基频特征与原始基频特征一样，则此处基频特征点重叠，图中只会显示原始音频基频，而不重叠部分即为攻击后出现偏差的基频。需要说明的是，这里提取出的基频特征并不一定属于音频同一声源的基频，选择最强基频是为了避免由于歌声歌唱间隙或乐器演奏停顿时，造成某一帧某一种声源基频特征缺失从而导致无法与零水印图像进行运算的情况，所以本文设计的方案提取出的基频特征在某些帧会出现图 9 中显示落差较大的情况。

由图 4(a)和(b)可以看出上、下采样对基频特征完全无影响；由(c)~(f)中可以看出低通滤波、噪声、MP3 压缩和抖动攻击后的基频出现偏差点，但是攻击后基频特征曲线与原始音频基频特征整体上基本一致。因此，由图 4 可以看出，在各种攻击下，音频的基频特征具有良好的稳定性。

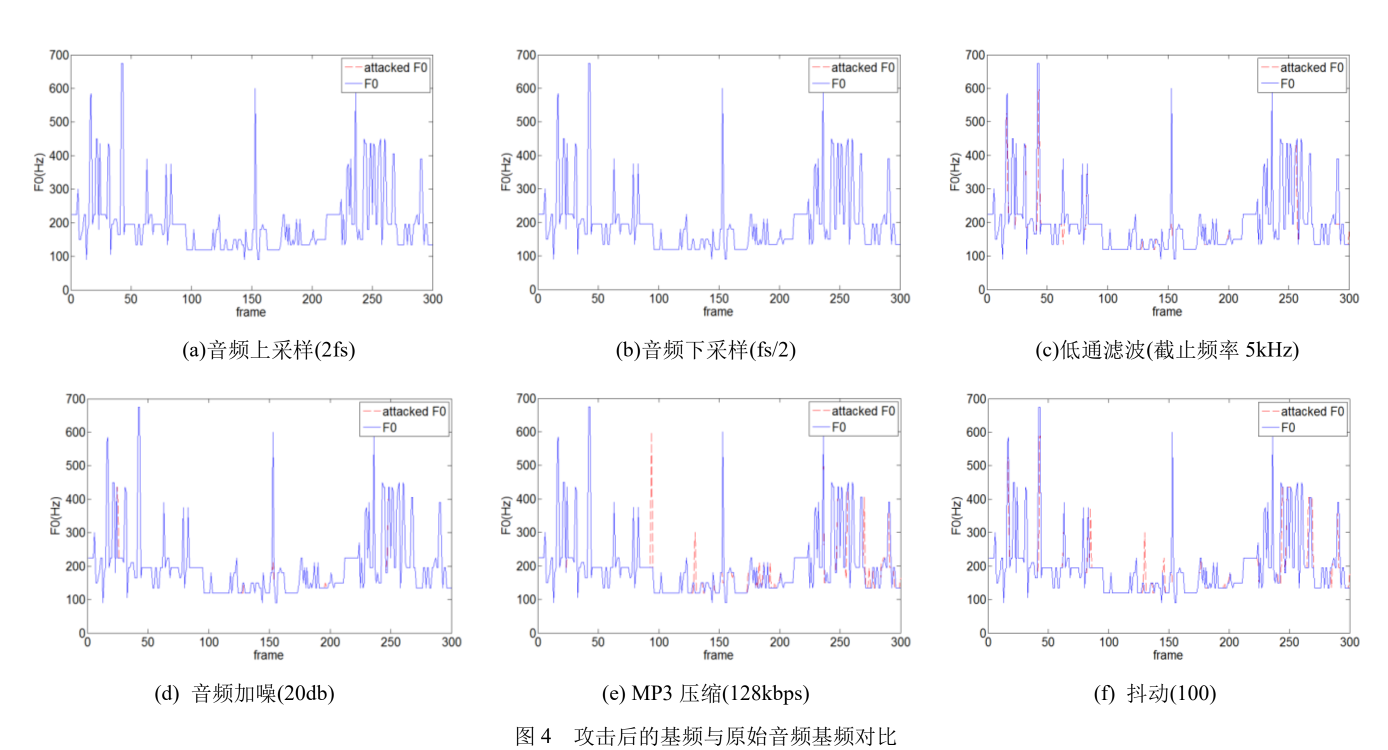


图 4　攻击后的基频与原始音频基频对比

此外，由于算法中每帧提取出的基频与其他帧的基频之间没有直接的关联，在遭到窜改攻击时，仅在窜改处的基频特征受到影响产生差异，而未受攻击处的基频特征并不会受到影响，所以利用该特征可以准确定位窜改。

3 水印嵌入和提取

3.1 基于基频特征生成零水印二值序列

图 5 所示的是零水印二值序列生成，即零水印的嵌入过程。

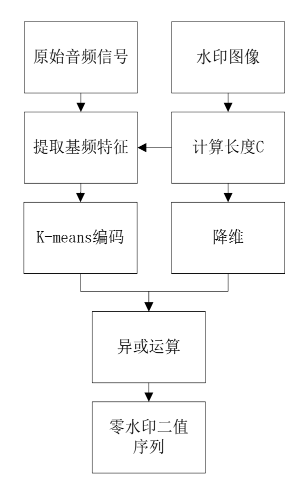


图 5　零水印二值序列生成过程

以大小为 P×Q 的有意义的二值图像作为本文方法的水印图像，其中 P 和 Q 为水印图像像素的行数和列数，P×Q=n，n 为总像素点数，水印图像可表示为 $i=\{w(a,b), 0\le a\le P, 0\le b\le Q\}$。零水印序列的生成步骤如下：

a)将原始二维的水印图像降维成一维序列，则此二值图可表示为

$$I=\{I(i), 0\le i\le n\} \quad (2)$$

b)记录已提取的基频特征为 f(i)，其中 i 表示帧号，0≤i≤n，n 表示总帧数。为了表达基频特征变化的整体趋势特征，本文选择 0、1 编码来表示基频特征，这就需要将特征分为两类，将其中一类编码为 1，另一类编码为 0，采用 k-means 聚类方法可以快速将基频特征 f(i)分成两类 C1 和 C2，然后根据式(3)将其编码成一个特征二值序列 K(i)。

$$K(i)=\begin{cases}1, & f(i)\in C1\\ 0, & f(i)\in C2\end{cases} \quad (3)$$

c)对水印图像 I 与特征二值序列 K(i)进行异或运算，即可得到含有特征的二值序列 W。W 为

$$W(i)=I(i)\oplus K(i) \quad (4)$$

d)记录并保存得到的零水印二值序列 W。

3.2　零水印图像提取

图 6 为零水印图像的提取过程。

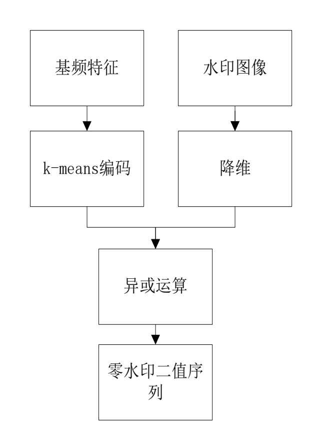


图 6　零水印提取过程

零水印的提取不需要原始音频载体，只需要基频特征序列 K 和二值水印序列 W，具体步骤与生成二值序列相似。

a)根据各种攻击后的音频信号得到的加攻击后的基频特征 f'(i)，0≤i≤n,其中 i 为帧号，n 为总帧数，使用 K-means 聚类得到基频特征的二值序列 K'(i)。

b)将特征二值序列 K'(i)与之前得到的零水印二值序列 W 异或，即可提取出水印图像 I'(i), I'(i)为

$$I'(i)=W(i)\oplus K'(i) \quad (5)$$

4　仿真实验及结果分析

实验部分分为鲁棒性测试和窜改检测两部分，分别对零水印算法的鲁棒性和窜改检测功能进行实验和结果分析。

4.1　鲁棒性测试

实验使用 MATLAB 2010b 软件进行仿真测试，对不同类型测试音频信号进行上采样（速度为 2 倍原始采样频率 fs）、下采样（速度为原始采样频率 fs 的 1/2）、低通滤波（截止频率为 5 kHz）、加高斯白噪声（20 dB）、 MP3 压缩（速率为 128 kbps）、抖动攻击（100）。采用归一化相关系数(NC)和比特误码率(BER)作为测试指标进行鲁棒性测试及分析。水印图选用一幅有意义的二值图像，大小为 32×32。

图 7 是在流行乐测试样例下，音频的采样频率(fs)为 44 100 Hz，时间长度为 27 s，在各种攻击下，本算法提取出的水印图像的结果。

图 7　各种攻击下提取出的水印图像

由图 7 中提取出的水印图像可以看出，本文提出的零水印算法在不同攻击下，仍然可以提取出清晰的水印图像，证明了基于基频特征的零水印算法在以上常见的各种攻击下具有较强的稳定性。

对于受到各种攻击后提取出的水印图像，计算水印算法的两个鲁棒性评价指标 NC 值和 BER 值，其中，NC 值越大，BER 值越小，则说明受到攻击后提取出的水印图像与原水印图像越接近，水印算法的抵抗各种攻击的鲁棒性越强。实验选取了流行乐、轻音乐和摇滚乐三种不同风格的音乐，并针对每种风格的音乐选取同一段音频，与文献[3]和文献[4]的算法进行鲁棒性对比测试，结果如表 1~3 所示。

表 1　流行乐在各种攻击下提取水印的 NC 和 BER 值

攻击类型	本算法		文献[3]算法		文献[4]算法	
	NC	BER	NC	BER	NC	BER
上采样(2fs)	1	0	1	0	0.9977	0.0039
下采样(fs/2)	1	0	0.9994	0.0009	0.9994	0.0010
低通滤波	1	0	0.9930	0.0117	0.8821	0.1895
加噪 20db	0.9965	0.0059	0.9913	0.0146	0.9913	0.0146
MP3 压缩（128kbps）	0.9854	0.0244	0.9784	0.0361	0.9971	0.0049
抖动+100	0.9907	0.0156	0.9831	0.0283	0.9320	0.1113

表 2　轻音乐在各种攻击下提取水印的 NC 和 BER 值

攻击类型	本算法		文献[3]算法		文献[4]算法	
	NC	BER	NC	BER	NC	BER
上采样(2fs)	1	0	1	0	0.9988	0.0020
下采样(fs/2)	1	0	1	0	0.9994	0.0010
低通滤波	1	0	0.9930	0.0117	0.9959	0.0068
加噪 20db	0.9942	0.0098	0.9977	0.0146	0.9959	0.0068
MP3 压缩（128kbps）	0.9889	0.0186	0.6850	0.4570	1	0
抖动+100	0.9965	0.0059	0.9328	0.1104	0.9485	0.0850

表 3　摇滚乐在各种攻击下提取水印的 NC 和 BER 值

攻击类型	本算法		文献[3]算法		文献[4]算法	
	NC	BER	NC	BER	NC	BER
上采样(2fs)	1	0	1	0	0.9983	0.0029
下采样(fs/2)	1	0	0.9977	0.0039	0.9988	0.0020
低通滤波	0.9954	0.0078	0.9807	0.0322	0.9593	0.0674
加噪 20db	0.9907	0.0156	0.9820	0.0303	0.9913	0.0146
MP3 压缩（128kbps）	0.9930	0.0117	0.9954	0.0078	0.8249	0.2715
抖动+100	0.9742	0.0430	0.8682	0.2090	0.8347	0.2578

由表 1 实验结果可以看出，针对于流行音乐，本文提出的水印算法与文献[3]方案相比鲁棒性测试结果均优于其结果；在与文献[4]算法鲁棒性测试结果比较时发现，本算法在除 MP3 攻击外的其他常规攻击下均具有较强的鲁棒性，尤其在低通滤波和抖动攻击下明显优于文献[4]算法。由表 2 结果所示，对于轻音乐，本文提出的算法鲁棒性均良好，而文献[3]提出的水印算法在 MP3 压缩和抖动攻击下稳定性均比较差；文献[4]算法在抖动攻击下鲁棒性较弱。表 3 实验结果显示了对于摇滚乐，本文提出的算法在各种攻击下仍然很稳定，而文献[3]中的算法在抖动攻击下鲁棒性较差；文献[4]算法在 MP3 压缩和抖动攻击下稳定性均非常弱。这主要是由于基频特征具有较好稳定性，且本算法将提出的基频特征做 K-means 聚类编码，经编码生成的二值序列对各种攻击的敏感性降低，进一步增强了水印特征的稳定性。且本文中采用固定帧数有重叠的分帧方法，帧长随音频变化可调，音频在遭到攻击，尤其是抖动攻击时，局部变化引起时间尺度上的伸缩对于分帧后的音频信号的影响较小，文献[3]算法方案中采用的无重叠分帧方法是导致其抵抗抖动攻击性能不如本算法的一个原因；而文献[4]算法需要同时保证帧长为一个设定的值且帧数等于水印长度，当攻击导致音频发生变化时其每帧采样点发生较大改变，并且随着帧数增加这种偏差会变大，所以，对于测试的三种风格的音频，在抖动攻击下本文提出的零水印方案比文献[4]鲁棒性好。

综上所述，对于不同风格的音乐，本文提出的基于基频特征的零水印算法不仅在各种普通攻击及抖动攻击下均具有良好的鲁棒性，而且基频特征相较于其他文献中提取的系数特征或符号特征对于音频内容而言具有更为实际和直观的意义。

4.2　窜改检测定位

根据本文提出的算法每帧音频基频特征是互无关联地提取的特点，如果音频受到恶意窜改，仅被窜改部分音频的基频特征会发生改变，因此通过比较音频的基频特征即可确认是否发生窜改，若某些帧的基频特征出现不一致且这些帧集中在一定的区域，则说明音频在此区域内受到了局部窜改，定位此区域为窜改区域。

图 8(上)为原始音频信号波形图，图 8(中)是对原始音频的 5~6 s 静音窜改后的波形图，图 8(下)为原始音频的 2~3 s 被 5~6 s 处替换后的音频波形，为突出展示窜改部分，选取音频的前 6 s 展示在图片中。

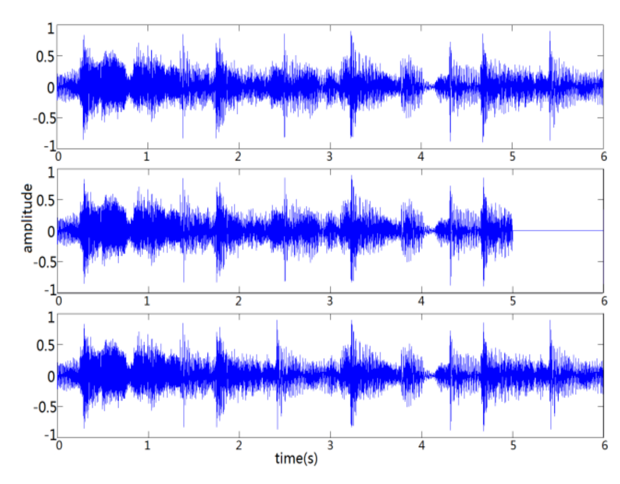
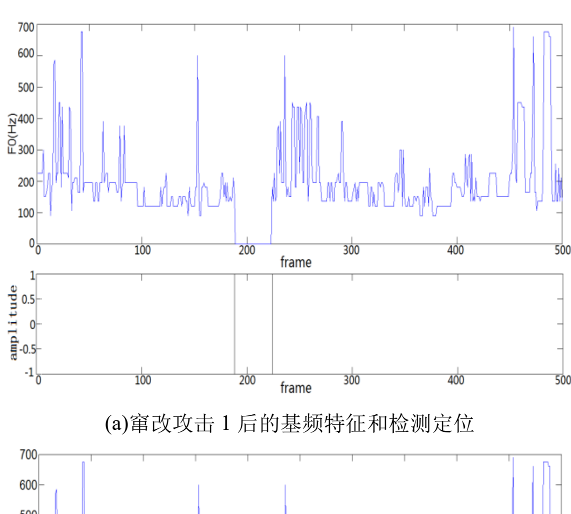
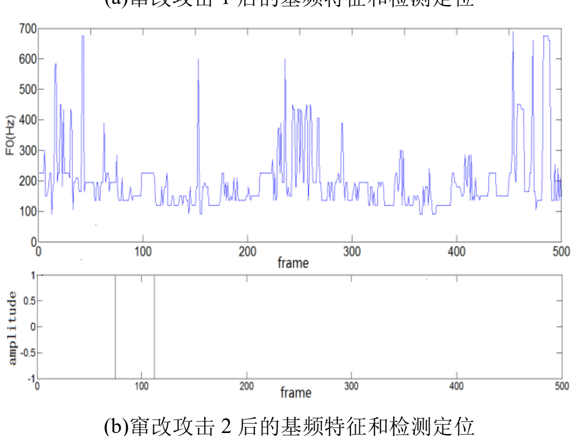


图 8　原始音频信号和窜改后的音频信号

窜改攻击后的基频特征及检测定位结果如图 9 所示。为放大窜改的部分以便观察，选取前 500 帧的结果在图片中展示。为了清楚地对比展示窜改检测的结果，图 9 中将窜改后的基频特征和检测定位放在同一幅图中，图 9(a) 中上半部分的特征图是将第 5~6 s（约 188~224 帧）静音处理后的音频的基频特征，图 9(a) 下半图为静音窜改后在音频信号上的检测定位结果，图 9(b)是用 5~6 s（约 188~224 帧）替换 2~3 s（76~112 帧）处的音频特征(上)及替换后在音频信号上窜改检测定位的结果(下)。



(a)窜改攻击 1 后的基频特征和检测定位



(b)窜改攻击 2 后的基频特征和检测定位

图 9　窜改后的音频特征及定位

由图 9 (a)中显示结果可以看出，音频特征在受到恶意攻击 1 时，188~224 帧的基频丢失，即 188~224 帧处被窜改出现静音，可以看出静音部分被成功在音频信号上定位。图 9 (b)的实验结果显示，音频特征在受到恶意攻击 2 时，76~114 帧音频信号被窜改；同时还可以看出，窜改部分被准确定位。综上所述，本文设计的算法在恶意窜改攻击下，能够较为准确地检测并定位音频信号被攻击的位置。

5　结束语

本文算法是在音频内容分析的基础上，对采用 NSHS 算法提取出的具有代表音频内容信息的基频特征，进行 K-means 聚类并编码为一组长度与降维后的水印图像长度一致的 0-1 序列，然后与水印图像异或，从而构造出二值水印序列。采用零水印算法没有改变音频的内容，保证了算法具有良好的隐蔽性。实验证明，由于音频基频特征具有良好的稳定性，所以由基频特征构造出的二值序列对各种攻击具有良好的鲁棒性。此外，提取出的基频特征彼此间没有直接关联，在恶意窜改攻击下，仅窜改处的特征受到影响，所以由实验结果看出，本方案可以达到窜改检测的目的。

参考文献：

[1] Kumar S, Dutta A. Performance analysis of spatial domain digital watermarking techniques [C]// Proc of IEEE International Conference on Information Communication and Embedded Systems. 2016: 1-4.

[2] 高华玲. 信息隐藏关键技术研究综述 [J]. 电子世界, 2016 (9): 146-148.

[3] 孙瑞鹏, 徐海涛. 基于 DCT 系数符号的音频零水印算法 [J]. 计算机技术与发展, 2014, 24 (5): 146-149.

[4] 蔡咏梅, 郭文强. 基于 DWT-SVD 的音频零水印算法 [J]. 计算机工程与设计, 2014, 35 (1): 42-46.

[5] Gopalan K, Fu J. An imperceptible and robust audio steganography employing bit modification [C]// Proc of IEEE International Conference on Industrial Technology. 2015: 1635-1638.

[6] 杨得国, 李智, 姜金娣. 基于音频特征和低频系数较小值的水印算法 [J]. 计算机工程, 2012, 38 (21): 10-13.

[7] 刘光玉, 张雪英, 马朝阳. 用于内容认证的半脆弱音频零水印算法 [J]. 计算机应用, 2012, 32 (4): 976-980.

[8] Wu Weina. Digital audio blind watermarking algorithm based on audio characteristic and scrambling encryption [C]// Advanced Information Technology, Electronic and Automation Control Conference. 2017: 1195-1199.

[9] 杨榆, 雷敏, 程明智, 等. 一种基于能量比较的音频零水印方案 (英文) [J]. 中国通信, 2014, 11 (7): 110-116.

[10] 张雪源. 面向音频检索的音频特征分析方法研究 [D]. 广州: 华南理工大学, 2015.

[11] Cheng T, Xu W, Tian Y, et al. Extracting singing melody in music with accompaniment based on harmonic peak and subharmonic summation [C]// Proc of Iet International Conference on Wireless, Mobile & Multimedia Networks. 2012: 200-205.

[12] Hsu C L, Chen L Y, Jang J S R, et al. Singing pitch extraction from monaural polyphonic songs by contextual audio modeling and singing harmonic enhancement [C]// Proc of International Society for Music Information Retrieval Conference. 2009: 201-206.

[13] Ikemiya Y, Itoyama K, Yoshii K, et al. Singing voice separation and vocal F0 estimation based on mutual combination of robust principal component analysis and subharmonic summation [J]. IEEE//ACM Trans on Audio Speech & Language Processing, 2016, 24 (11): 2084-2095.

[14] 宋岳阳. 基于单源欠定语音分离的音乐主旋律提取方法研究 [D]. 北京: 北京邮电大学, 2012.